Carlsberg Res. Commun. Vol. 48, p. 529–544, 1983

THE COMPLETE AMINO ACID SEQUENCE OF THE GLYCOPROTEIN, GLUCOAMYLASE G1, FROM ASPERGILLUS NIGER

by

BIRTE SVENSSON, KJELD LARSEN[1)]

and

IB SVENDSEN

Carlsberg Laboratory, Department of Chemistry,
Gamle Carlsberg Vej 10, DK-2500 Copenhagen Valby

and

ESPER BOEL

Laboratory of Genetics,
NOVO Research Institute, Novo Allé, DK-2880 Bagsværd

[1)] Present address: Harvard Medical School,
Center for Biochemical and Biophysical Sciences and Medicine,
Seeley G. Mudd Bldg., 250 Longwood Ave., Boston, MA 02115

Keywords: Peptide sequencing, cyanogen bromide, hydroxylamine, oxidative halogenation, S. aureus V8 protease, trypsin, endoproteinase Lys-C, deglycosylation, nucleotide sequencing

The primary structure of glucoamylase G1 (EC 3.2.1.3) from Aspergillus niger has been determined. Fragments of G1 were obtained by cleavage with either cyanogen bromide, hydroxylamine, or S. aureus V8 protease. The resulting peptides were separated using ion exchange chromatography on DEAE-Sephacel, gel filtration, and affinity chromatography on Con A-Sepharose. Secondary fragments were generated by cleavage with either o-iodosobenzoic acid or BNPS-skatole as well as by digestion with S. aureus V8 protease, trypsin, and endoproteinase Lys-C. These peptides were purified by the procedures mentioned above and by reverse phase HPLC. The present fragments were amino acid sequenced and this permitted, in combination with the tryptic peptides (Carlsberg Res. Commun. 48, 517-527 (1983)), identification of 574 of the 614 amino acid residues in G1. Sequencing of glucoamylase G1 cDNA, constructed from A. niger total mRNA, enabled deduction of the sequence of the remaining 40 amino acid residues localized to 6 short stretches. From the alignment of the fragments the complete primary structure of the enzyme was established. The amino acid sequence corresponds to a molecular weight of the polypeptide moiety of 65,424. Including both hexosamine and neutral carbohydrate contents the molecular weight of the present sample of G1 was calculated to be about 82,000.

The majority of the carbohydrate of G1 is found in a highly glycosylated region of 70 amino acid residues which comprises about 35 O-glycosyl serine and threonine residues. This region ends approximately 100 residues from the C-terminus of the enzyme. Two N-glycosylated positions were found in the central part of the polypeptide chain. The molecule contains 9 half-cystine residues. No homology is apparent between the sequence of glycoamylase and various α-amylases.

Abbreviations: BNPS-skatole = 2-(2-nitrophenylsulfenyl)-3-methyl-3'-bromoindolenine; ca = citraconyl; Con A = concanavalin A; DFP = diisopropylfluorophosphate; DPCC = diphenylcarbamyl chloride; EDTA = ethylenediaminetetraacetic acid, disodium salt; Hepes = N-2-hydroxyethyl piperazine-N'-2-ethanesulfonic acid; Nemac = N-ethylmorpholine acetate; HPLC = high pressure liquid chromatography; PTH = phenylthiohydantoin; SDS-PAGE = polyacrylamide gel electrophoresis in the presence of sodium dodecyl sulfate; 2-pe = 2-pyridylethyl; G1 designates the larger and G2 the smaller of the forms of glucoamylase from A. niger (44).

Springer-Verlag 0105-1938/83/0048/0529/$03.20

1. INTRODUCTION

Glucoamylases (1,4-α-D-glucan glucohydrolase, EC 3.2.1.3) are industrially important enzymes, which catalyze the release of D-glucose from the non-reducing ends of starch and related oligo- and polysaccharides (32, 35). The glucoamylases are produced by a wide variety of microorganisms (15, 17, 28, 32, 45, 48). They frequently exist in multiple forms of which only the larger, referred to as G1, has the capacity to adsorb to and digest raw starch (33, 44, 49). Partial amino acid sequence analysis of G1 and the smaller form G2 from A. niger suggested that the two forms have highly homologous structures, although G1 is extended with a C-terminal fragment that is lacking in G2 (44).

The glucoamylase from A. niger is an unusual glycoprotein. It is rich in mannose and carries a large number of short sugar units linked to serine and threonine (33, 34, 44) which are mostly located in a highly glycosylated region comprising 70 amino acid residues (43). Evidence for a few additional attachment sites for carbohydrate outside this region has been obtained; among these are two N-glycosylated asparagine residues.

In view of the wide commercial interest in glucoamylases and the future usefulness in studies of their structure-function relationships, we report the completion of the amino acid sequence of glucoamylase G1 from A. niger which has been achieved by a combination of results from peptide and cDNA sequencing.

2. MATERIALS AND METHODS

2.1. Materials

Homogeneous glucoamylase G1 from Aspergillus niger was prepared as described by SVENSSON et al. (44) from AMG 200L (Batch ASN 5192), a commercial preparation of the enzyme produced by Novo Industries, Bagsværd, Denmark.

Bio-Gels P-150, P-100, P-60, P-30, P-10, P-6, and P-2 were from Bio-Rad, Richmond, CA. DEAE-Sephacel and Con A-Sepharose were products of Pharmacia Fine Chemicals, Uppsala, Sweden. Dithiothreitol, DPCC-treated trypsin, DFP-treated carboxypeptidase B, cyanogen bromide, tryptamine hydrochloride, and nuclease S1 (EC 3.1.30.1; from Aspergillus oryzae) were from Sigma, St. Louis, MO. S. aureus V8 protease was from Miles, Stoke Poges, U.K.; endoproteinase Lys-C was from Boehringer, Mannheim, F.R.G.; 2-vinylpyridine was from Aldrich-Europe, Beerse, Belgium; o-iodosobenzoic acid and BNPS-skatole were from Pierce, Rockford, IL. Hydroxylamine hydrochloride, citraconic anhydride and tyramine hydrochloride were obtained from Merck, Darmstadt, F.R.G.. Trifluoroacetic acid was from Rathburn, Walkerburn, Scotland, as were reagents and solvents for the sequencer except ethylacetate which was from Merck. Carboxypeptidase Y was prepared in this laboratory. Spectrapor dialysis tubings with cutoff at molecular weights of 3,500 and 6,000-8,000 were from Spectrum medical Instruments, Los Angeles, CA. DNA polymerase I "Klenow fragment" (EC 2.7.7.7; from E. coli), and oligo(dT)$_{12-18}$ were purchased from P-L Biochemicals, Milwaukee, WI. Reverse transcriptase (RNA-dependent DNA polymerase, EC 2.7.7.4; from avian myeloblastosis virus) was purchased from J.W. Beard, Life Sciences, St. Petersbourg, FL. Terminal deoxynucleotidyl transferase (EC 2.7.7.31) was from Bethesda Research Laboratories, Gaithersburg, MD. Restriction endonucleases were from New England BioLabs., Beverly, MA. RNasin (human placenta ribonuclease inhibitor) was from Biotec, Madison, WI. Oligo(dT)-cellulose was type T-2 from Collaborative Research, Waltham, MA. [α-^{32}P]-dCTP (3,200 Ci/mmol) and [α-^{32}P]-dATP (3,200 Ci/mmol) were obtained from New England Nuclear.

2.2. Methods

2.2.1. Chemical modifications

Glucoamylase G1 was dissolved (60 mg · ml^{-1}) in 7 M-guanidine hydrochloride, 0.2 M-Tris-HCl and 5 mM-EDTA pH 8.3 and treated with dithiothreitol followed by 2-vinylpyridine to yield 2-pyridylethylated G1, as earlier reported (12, 43). Citraconylation was performed a.m. DIXON and PERHAM and deglycosylation of peptides a.m. MORT and LAMPORT (5, 30, 43).

2.2.2. Procedures for cleavage of peptide bonds

In order to ensure that the methionine residues were in the reduced state prior to cyanogen

bromide cleavage, G1 was dissolved in the guanidine hydrochloride/Tris-HCl buffer (section *2.2.1*) and treated with 5% (vol/vol) 2-mercaptoethanol for 24 hours at room temperature (19). The solution was then dialyzed against 10% (vol/vol) ethanol, containing 0.1% 2-mercaptoethanol, and lyophilized. The reduced G1 was dissolved (50 mg · ml^{-1}) in 70% (vol/vol) trifluoroacetic acid containing CNBr (50 mg · ml^{-1}) and tryptamine hydrochloride (7.5 mg · ml^{-1}). The reaction mixture was left for 20 hours, then 20 volumes of water was added and CNBr and solvent were removed by lyophilization. The cyanogen bromide fragments were reduced and alkylated (section *2.2.1*) and subsequently separated by DEAE-Sephacel chromatography (section *2.2.3*).

Cleavage with hydroxylamine was performed in principle as reported by BORNSTEIN and BALIAN (3). The freshly prepared reagent was 2 M-hydroxylamine in 6 M-guanidine hydrochloride, adjusted to pH 9.0 with 4.5 M-lithium hydroxide and equilibrated for 5 minutes at 45 °C. 2-pe-G1 was dissolved (2-5 mg · ml^{-1}) immediately in the reagent and incubated for 2-4 hours at 45 °C, during which period the pH was maintained at 9.0 by addition of lithium hydroxide. The reaction mixture was centrifuged, the supernatant obtained was passed through Bio-Gel P-2 in 0.05 M-ammonium bicarbonate to remove reagent and denaturant and the fragments were subsequently separated on Bio-Gel P-100.

Cleavage at the C-terminus of tryptophan residues was performed using either o-iodosobenzoic acid or BNPS-skatole according to procedures described by MAHONEY et al. (25) and FONTANA et al. (10), respectively. All reactions were carried out in the dark at room temperature. o-Iodosobenzoic acid (20 mg · ml^{-1}) and tyramine hydrochloride (4 mg · ml^{-1}) in 4 M-guanidine hydrochloride and 80% in acetic acid were preincubated for 6 hours. The separated cyanogen bromide fragments were dissolved (2-5 mg · ml^{-1}) in the reagent and the solution was left for 22 hours. The reaction was then stopped by gel filtration on Bio-Gel P-6 in 30% acetic acid. For each of the cyanogen bromide fragments the two pools of peptides obtained were lyophilized. The C-terminal cyanogen bromide fragment (2-8 mg · ml^{-1}) was also cleaved by BNPS-skatole (82 mg · ml^{-1}) in 4 M-guanidine hydrochloride and 80% in acetic acid in the presence of tyrosine (20 mg · ml^{-1}). After 18 hours with stirring 20 volumes of water were added to the reaction mixture and BNPS-skatole was removed by extraction with diethylether. The aqueous phase was adjusted to pH 7.0 with NH_3, aq., and passed through Bio-Gel P-2 in 0.1 M-ammonium bicarbonate. The pool of peptides was lyophilized.

The N-terminal cyanogen bromide fragment (1 mg · ml^{-1}) was digested with endoproteinase Lys-C (1:150, wt/wt; 16 hours at 37 °C) in 0.06 M-ammonium bicarbonate pH 7.7 (38). The C-terminal cyanogen bromide fragment (4 mg · ml^{-1}) was digested with S. aureus V8 protease (1:40, wt/wt; 3 hours at 40 °C) in 0.1 M-ammonium bicarbonate pH 7.8. Digestion of 2-pe-G1 (1.2 mg · ml^{-1}) with the same enzyme (1:40, wt/wt; 12 hours at 40 °C) was carried out in 0.07 M-sodium phosphate pH 7.6 to achieve cleavage after both glutamic acid and aspartic acid residues (16). Peptide fragments (2 mg · ml^{-1}) were digested with trypsin (1:50, wt/wt; 2-4 hours at 37 °C) in 0.1 M-ammonium bicarbonate pH 7.8, 0.1 mM-calcium chloride.

2.2.3. Separation of peptides

Cyanogen bromide fragments, some of their subfragments, and selected hydroxylamine fragments were purified on DEAE-Sephacel (1.6×60 cm) which was equilibrated in 7 M-urea, 0.05 M-Tris-HCl pH 8.0. Elution was performed with this buffer (200 ml) followed by a linear gradient from zero to 0.2 M-guanidine hydrochloride in the same buffer (2×600 ml). Pools of peptides were desalted by dialysis against 0.05 M-ammonium bicarbonate using Spectrapor dialysis tubing and lyophilized. Smaller peptides were desalted by gel filtration through Bio-Gel P-6 in either 0.05 M-ammonium bicarbonate or 30% acetic acid.

Separation of peptides (1-15 mg) by gel filtration was performed through either Bio-Gels P-150, P-100, P-60, P-30, P-10, or P-6 (1.5×90 cm) in 0.2 M-ammonium bicarbonate pH 7.8, unless otherwise specified.

Glycopeptides, generated by digestion with S. aureus V8 protease or cleavage with BNPS-skatole, were purified by affinity chromatogra-

phy on Con A-Sepharose in principle as earlier described (43).

A Waters chromatograph was used for purification of peptides by reverse phase HPLC on a Wide Pore C_{18} column, i.d. 4.6 mm and length 25 cm, from J.T. Baker Research Products (43). Elution was performed with linear gradients from 0.1% trifluoroacetic acid to 0.1% trifluoroacetic acid in 80% 1- or 2-propanol (24).

2.2.4. Amino acid sequence analysis

Automated sequencing was performed with a Beckman Sequencer 890C and PTH-amino acids were identified as recently described (8, 18, 20, 41, 43). PTH-2-pe-Cys was identified in the water phase. After oxidative halogenation PTH-Tyr coeluted with PTH-Val by HPLC. Digestion with carboxypeptidase B (0.02 mM) was performed at a molar ratio from 1:100 to 1:500 in 0.05 M-Nemac pH 8.5 at room temperature or in 0.05 M-Hepes pH 7.5 at 40 °C. Carboxypeptidase Y (0.02 mM) in 0.1 M-Nemac pH 5.0 at a molar ratio from 1:100 to 1:1000 was used either alone or after carboxypeptidase B. Aliquots were removed at appropriate time intervals and the amino acids released were determined.

2.2.5. Preparation of mRNA

Aspergillus niger cells growing at conditions optimized for glucoamylase production were pelleted at 10,000×g and 4 °C for 10 minutes. The pelleted cells were frozen in liquid nitrogen and stored at +80 °C until RNA extraction.

To extract the RNA, 5 g of frozen cells were pulverized under liquid nitrogen in a mortar. Quartz was added to disrupt the cell wall. The resulting powder was suspended in 60 ml of a buffer containing 5 M-guanidinium rhodanide, 50 mM-Tris-HCl at pH 7.5, 10 mM-EDTA, 5% (vol/vol) 2-mercaptoethanol, 4% (wt/vol) N-lauroylsarcosine, and 15% (wt/vol) CsCl. A few strokes with an Ultra Turrax homogenizer was used to decrease the viscosity of the suspension by shortening the length of chromosomal DNA from the cells. After pelleting insoluble material at 10,000×g at 4 °C for 15 minutes, total RNA was isolated from the supernatant as described (4) by pelleting through a 5.7 M-CsCl/100 mM-EDTA cushion in an SW41 Beckman rotor run at 37,000 rpm for 18 hours at room temperature. Poly(A)-containing RNA was obtained by two cycles of chromatography on oligo(dT)-cellulose (1).

2.2.6. Construction and cloning of double-stranded glucoamylase G1 cDNA

Double-stranded Aspergillus niger cDNA was synthesized and cloned on pBR327 in E. coli by procedures as described (2). Details of the cloning and the identification of glucoamylase G1 specific recombinants in the resulting cDNA library will be presented elsewhere.

2.2.7. Analytical procedures

The amino acid composition, contents of O-glycosylated serine and threonine residues, glucosamine, and neutral carbohydrate were determined as previously reported (6, 7, 43). SDS-PAGE was carried out for selected large fragments a.m. LAEMMLI (22) using 15% acrylamide gels as earlier described (42). The DNA sequence of the inserts in glucoamylase G1 specific recombinants was determined by the chemical cleavage procedure (27).

2.2.8. Peptide nomenclature

Fragments obtained by cleavage of glucoamylase G1 with cyanogen bromide are named CB fragments and designated CB1-CB3, starting from the N-terminus of the protein. The hydroxylamine fragments (Hy) and S. aureus V8 protease fragments (E) from G1 are similarly numbered.

Secondary fragments purified after either cleavage at tryptophan residues (W), or digestions with trypsin (T), with trypsin after citraconylation (Tca), with endoproteinase Lys-C (K), or with S. aureus V8 protease (E) are named according to their position in the primary fragment, viz. W2-CB1 is the second from the N-terminus of the isolated tryptophan peptides from CB1. The numbering of the tryptic fragments prepared from 2-pe-G1 and citraconylated 2-pe-G1 is as published (43). Thus of the sequenced peptides found in pool 6 from the gel filtration through Bio-Gel P-100, T6-3 was the third to be released from the HPLC column (43). Localization of the fragments in the G1 sequence is often indicated by a parenthesis with the numbers of their N- and C-terminal residues.

3. RESULTS

3.1. Cyanogen bromide fragments

The glucoamylase G1 from A. niger contains two methionines. When the enzyme was cleaved with cyanogen bromide and the mixture of fragments was sequenced, the N-termini of CB1, CB2, and CB3 were found to be present in yields of 100%, 82%, and 70%, respectively. This improvement, compared with the previous results (44), is explained by the trifluoroacetic acid effecting better cleavage than the weaker formic acid of the Met-Ser bond (39). The 2-pyridylethylated CB fragments were separated by ion exchange chromatography on DEAE-Sephacel (Figure 1). Pool A contained CB1, while CB2 and CB3 were found in pools D and B, respectively. Pool C contained incompletely cleaved fragments. Materials from the different peaks within pool A (Figure 1) had essentially the same amino acid composition suggesting charge heterogeneity within CB1. Consistent with this, both aspartic acid and asparagine were found by the automated sequencing at positions 20 and 21 from the N-terminus of CB1.

The amino acid composition, contents of O-glycosylated serine and threonine, glucosamine, and neutral carbohydrate were determined after rechromatography of the CB fragments (Table I) and agree with the results found for intact G1. Deglycosylation of G1 by HF at 0 °C for 3 hours removed one glucosamine residue, suggested to be the residue found in CB3. From the absence of homoserine and homoserine lactone in CB3 and the identical N-terminal sequences of CB1 and G1 (44), the order of the fragments in the primary structure was established to be CB1(1-133)-CB2(134-396)-CB3(397-614) (Figure 2). In Figure 2 (after page 542) are also aligned tryptic fragments (43) and the fragments discussed below, which enabled the reconstruction of the amino acid sequence of G1.

3.2. Hydroxylamine fragments

The hydroxylamine fragments of G1 were separated by gel filtration on Bio-Gel P-100 (Figure 3). The subsequent steps of purification and the positions of the purified peptides in the sequence are listed in Table II. Hy1-Hy4 were generated by cleavage of Asn-Gly bonds, and Hy5 by the exceptional cleavage of the Asn-Ser bond found

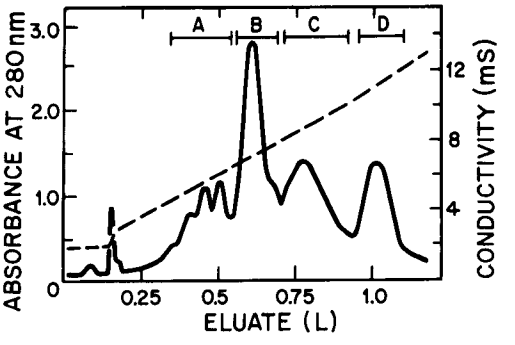

Figure 1. Separation of cyanogen bromide fragments (260 mg) on DEAE-Sephacel (see *2.2.3*). Fractions of 11 ml were collected, pooled as indicated and subsequently rechromatographed in the same system. The CB fragments are described in Table I.

within Hy4. The apparent molecular weights of Hy1, Hy3, and Hy4 were estimated by SDS-PAGE to be 11,000, 16-19,000, and 44,000, respectively. The elution position of Hy4 from Bio-Gel P-100 (Figure 3; Table II) suggested a folded conformation for this long fragment representing the C-terminal half of the polypeptide chain of G1.

3.3. S. aureus V8 protease fragments

2-Pyridylethylated G1 was digested with S. aureus V8 protease in the presence of phosphate. The resulting glycopeptides were isolated by Con

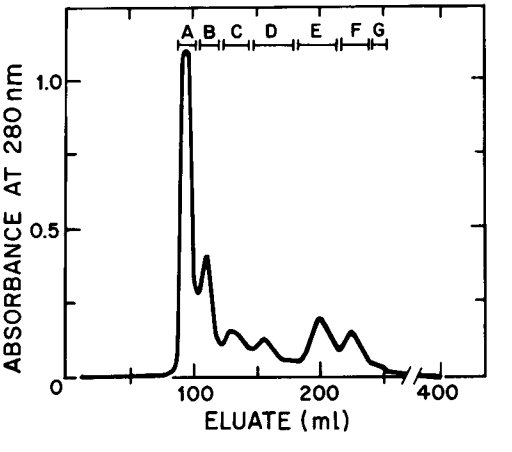

Figure 3. Separation of hydroxylamine fragments (21 mg) on Bio-Gel P-100 (2.5×90 cm). The indicated pools, A, B, C, E, and F were further purified to give the fragments described in Table II.

Table I.
Chemical characterization of cyanogen bromide fragments of glucoamylase G1 from Aspergillus niger.

Amino acid	Residues/fragment						Residues/protein		
	CB1		CB2		CB3		CB(1+2+3)	G1	
Aspartic acid	18.4	(18)	30.5	(31)	18.6	(19)	68	67.3	(67)
Threonine[a]	11.6	(12)	19.6	(20)	40.6	(41)	73	72.5	(73)
Serine[a]	15.1	(15)	33.8	(34)	33.9	(34)	83	83.6	(84)
Glutamic acid	7.4	(7)	21.1	(21)	15.0	(15)	43	43.2	(43)
Proline	6.0	(6)	7.4	(7)	8.9	(9)	22	22.9	(23)
Glycine	13.4	(13)	17.4	(17)	15.2	(15)	45	45.3	(45)
Alanine	14.0	(14)	26.7	(27)	21.0	(21)	62	62.0	(62)
Valine[b]	7.8	(8)	16.8	(17)	14.6	(15)	40	40.3	(40)
Methionine								2.2	(2)
Isoleucine[b]	6.7	(7)	9.0	(9)	7.1	(7)	23	22.9	(23)
Leucine	11.5	(12)	21.0	(21)	11.2	(11)	44	45.1	(45)
Tyrosine	4.3	(4)	14.3	(14)	9.0	(9)	27	27.0	(27)
Phenylalanine	3.2	(3)	15.0	(15)	3.9	(4)	22	21.9	(22)
Histidine	0.2	(0)	3.8	(4)	0.3	(0)	4	4.2	(4)
2-pe-Cysteine[c]	0.2	(0)	4.6	(5)[f]	2.5	(3)	8	7.9	(8)
Lysine	1.8	(2)	4.6	(5)	4.5	(5)	12	11.5	(12)
Arginine	5.7	(6)	6.3	(6)	5.5	(6)	18	18.4	(18)
Homoserine[d]	0.7	(1)	0.8	(1)	n.d.		2		
Tryptophan[e]	3	(3)	8	(8)	7	(7)	18	18	18
Total number of residues		131		262		221	614		616
O-Glycosylated threonine[g]	2.6	(3)	2.4	(2)	21.1	(21)			
O-Glucosylated serine[g]	0.7	(1)	4.2	(4)	16.0	(16)			
Neutral carbohydrate[h]	5.3		21		73				
Glucosamine[i]	trace		4.0		1.0				

a) Extrapolated to zero time of hydrolysis
b) Value after 72 hours of hydrolysis
c) 2-Pyridylethyl-cysteine (12)
d) Sum of homoserine and homoserine lactone; n.d. means not detected
e) From the sequence determination
f) By sequencing, six 2-pe-Cys were found in CB2
g) Determined by reductive β-elimination followed by amino acid analysis (6)
h) Determined by the phenol-sulfuric acid procedure (7) using mannose as standard
i) Determined on the amino acid analyzer after acid hydrolysis (43)

The sizes of the CB fragments were based on: CB1, 6 Pro and 14 Ala; CB2, 21 Leu and 15 Phe; CB3, 15 Glu and 21 Ala. The apparent molecular weights were found by SDS-PAGE to be 40,000, 35,000, and 12,000 for CB2, CB3, and CB1, respectively.

A-Sepharose chromatography and subsequently separated by gel filtration through Bio-Gel P-60 which led to isolation of a large, glycosylated fragment, E1(438-518), with C-terminal aspartic acid. Digestion of E1 with carboxypeptidase Y was performed after deglycosylation. A dominating front peak containing E1-CB2(298-387, and in part 298-340) was obtained by separation of S. aureus V8 protease digests of CB2 on Bio-Gel P-100. The fragments obtained from CB3 by digestion with the S. aureus V8 protease were separated on DEAE-Sephacel (Figure 4). Two peptides were purified by gel filtration of pool A on Bio-Gel P-6. One was similar to E1 and the

Table II.
Purification of hydroxylamine fragments after separation on Bio-Gel P-100.

Pool (Figure 3)	Purification procedure	Fragment
A	Bio-Gel P-150	Hy3 (181-311)
B	DEAE-Sephacel, eluted at 0.07 M-GuHCl	Hy4 (312-614)
C	Bio-Gel P-60	Hy5 (429-614)
E	DEAE-Sephacel, eluted at 0.03 M-GuHCl followed by Bio-Gel P-60 in 0.02 M-GuHCl, 0.05 M-Tris-HCl, pH 7.6	Hy1 (68-143)
F	Bio-Gel P-100 followed by Bio-Gel P-60	Hy2 (144-180)

Chromatography on DEAE-Sephacel was followed by rechromatography in the same system. Gel filtration was performed as described in section *2.2.3* except in the case of Hy1. Hy5 is not included in Figure 2.

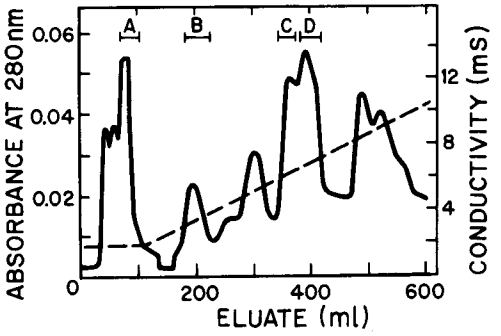

Figure 4. Separation of S. aureus V8 protease fragments from CB3 on DEAE-Sephacel (1.6×35 cm). Fractions of 5.5 ml were collected, pooled (indicated by bars) and further purified as described in 3.3.

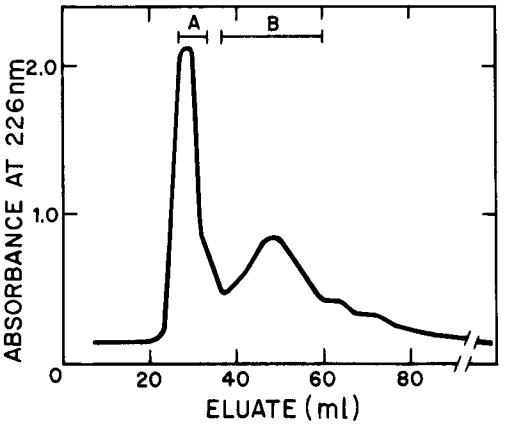

Figure 5. Separation on Bio-Gel P-60 of the larger fragments (4 mg) obtained by cleavage of CB2 with o-iodosobenzoic acid. The pools A and B were subjected to automated sequencing.

other was E3-CB3(596-614), the C-terminal fragment of the protein (Figure 2). Pool B contained the N-terminal fragment of CB3, Ser(397)-Glu(437). Pools C and D were further purified by gel filtration on Bio-Gel P-6 and P-10, respectively, to yield E1-CB3(528-542) and E2-CB3(543-574).

3.4. Tryptophan fragments

Fragments of CB1 obtained by treatment with o-iodosobenzoic acid were separated when the reagents were removed by gel filtration through Bio-Gel P-6 (see *2.2.2*). Automated sequencing of the front pool showed W1-CB1, starting at Thr(51), to be a major component. Thus cleavage had occurred after Tyr(50) (Figure 2) even in the presence of tyramine hydrochloride. The smaller peptides in the second pool were separated by HPLC; W2-CB1(119-133) eluted at 18% 1-propanol.

The larger peptides generated by cleavage of CB2 with o-iodosobenzoic acid were fractionated on Bio-Gel P-100 (Figure 5). Pool A contained W1-CB2, which could only be sequenced for a few cycles because of a sudden drop in repetitive yield at Asn(180), probably due to reactions with the peptide bond to the adjacent Gly(181) (Figure 2). Pool B contained W2-CB2(227-315).

The fragments obtained by o-iodosobenzoic acid cleavage of CB3 were separated by DEAE-Sephacel chromatography (Figure 6). Pools A, B, and C were passed through Bio-Gel P-30 to give

W2-CB3(465-541, with a significant content of 465-530), W1-CB3(435-464) (not shown in Figure 2), and W4-CB3(562-588), respectively. Both W1-CB3 and W2-CB3 were sequenced before and after deglycosylation. Fragments of CB3 were also prepared by cleavage with BNPS-skatole and separated on Bio-Gel P-30 (Figure 7). The pool A was passed through Con A-Sepharose. The non-adsorbed and the adsorbed fractions were subsequently fractionated on Bio-Gel P-30 to yield W3-CB3(542-561) and W5-CB3(589-613), which coeluted with equimolar amounts of W1-CB3(435-464). The type of bond between the two latter peptides was not identified, but a disulfide might have formed from Cys(447) and Cys(602) as a result of incomplete reduction and alkylation. When sequencing W5-CB3, no vacant positions were observed and this would indicate the absence of carbohydrate side chains. However, it cannot be excluded that the fragment was retained on Con A-Sepharose because some of the serine and threonine residues were partially glycosylated.

3.5. Digestions with endoproteinase Lys-C and trypsin

The digest of CB1 with endoproteinase Lys-C was fractionated on Bio-Gel P-60 (Figure 8). Pool A contained K1-CB1(60-106) and pool B K2-CB1 (107-133, and significant amounts of 108-133). A chymotrypsin-like specificity of an endoproteinase Lys-C preparation was previously observed (38). The same bond, Phe(107)-Asn(108), was also susceptible to trypsin. The only other example in G1 of trypsin catalyzing the hydrolysis at an atypical residue seemed to be in the case of fragment Tca2-1, terminated by Phe(517) (43).

In addition to the tryptic fragments from 2-pe-G1 and citraconylated 2-pe-G1 earlier described (43) and shown in Figure 2, a few tryptic subfragments, which have been useful for the reconstruction of the amino acid sequence, were prepared. Thus Hy3(181-311) was citraconylated and digested with trypsin. Fractionation by HPLC resulted in Tca1-Hy3(272-284) eluting at 16% 2-propanol. Digestion of CB2 with trypsin and isolation of the smaller peptides by HPLC resulted in T1-CB2(336-350), which eluted at 26% 1-propanol and had not been isolated from

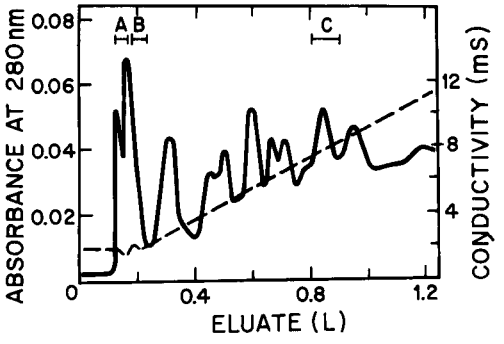

Figure 6. Separation on DEAE-Sephacel of fragments from CB3 obtained by cleavage with o-iodosobenzoic acid. Pools A, B, and C, were separately further purified on Bio-Gel P-30.

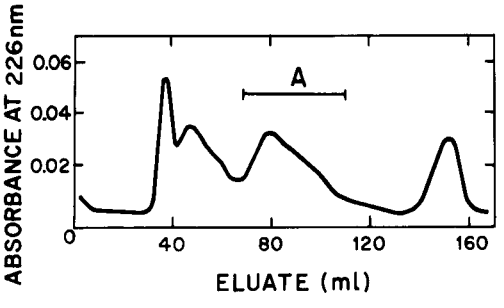

Figure 7. Separation on Bio-Gel P-30 of fragments from CB3 (2 mg) obtained by cleavage with BNPS-skatole. Pool A was further fractionated by affinity chromatography on Con A-Sepharose, described in 3.4.

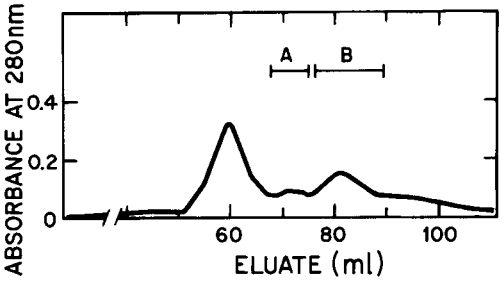

Figure 8. Separation on Bio-Gel P-60 of fragments (1.5 mg) from CB1 obtained by digestion with endoproteinase Lys-C. Pools A and B were subjected to automated sequencing.

tryptic digests of 2-pe-G1. Other isolated peptides were T6-2, T6-3 (243-271, not shown in Figure 2), T7-3, T8-3 described in an earlier publication (43), and T8-3+T7-3(272-284) which eluted

Table III.
Nucleotide sequences enabling the determination of the complete amino acid sequence of glucoamylase G1 from A. niger.

A. Amino acid residues identified only from the cDNA sequence data

UCU	CCG GCU CUG	AAC	CAC	GCA UGC GAC GAC UCC ACC UUC CAG CCC
Ser	Pro - Ala - Leu	Asn	His	Ala - Cys - Asp - Asp - Ser - Thr - Phe - Gln - Pro
93	126	169	191	259

AGU GAC AGC	ACU UAU AGU AGC AUU GUA GAU GCC GUG	AGC
Ser - Asp - Ser	Thr - Tyr - Ser - Ser - Ile - Val - Asp - Ala - Val	Ser
294	367	489

CCC ACC GCC	CUG ACA GCU ACC ACC ACC UAC GGC GAG
Pro - Thr - Ala	Leu - Thr - Ala - Thr - Thr - Thr - Tyr - Gly - Glu
510	519

B. Confirmation of non-overlapping peptide sequences

CCG CGC GCG CUC	UUC CGC UCA AUC	GUG AAG ACU UUC	AGC AAG ACC AGC
Pro - Arg - Ala - Leu	Phe - Arg - Ser - Ile	Val - Lys - Thr - Phe	Ser - Lys - Thr - Ser
270	283	375	496

UUC GAU CUG ACA	UAC GGC GAG AAC
Phe - Asp - Leu - Thr	Tyr - Gly - Glu - Asn
517	525

at 14% 1-propanol. Digestion of T7-3(278-284) by carboxypeptidases B and Y in combination released stoichiometric amounts of Arg, Phe and Asp. However, a component eluting between Thr and Asp in the chromatogram of amino acids, suggested the serine residue which was identified by amino acid analysis after acid hydrolysis of T7-3, to be modified. A tryptic subfragment of Hy2, Asn(159)-Asn(180), was eluted at 33% 1-propanol. It contained 1.6 residues of glucosamine, probably bound to the Asn(169) which has been identified by nucleotide sequencing (Table III). The highly glycosylated fragment, Tca2-1 (43) was digested with carboxypeptidase Y after deglycosylation resulting in identification of the five C-terminal residues.

3.6. Molecular cloning of glucoamylase specific cDNA

An oligo(dT) primed cDNA library synthesized from A. niger poly(A)-containing RNA was constructed and screened for glucoamylase specific recombinants. DNA restriction fragments were generated from clones with a cDNA complementary to glucoamylase mRNA, and labeled with radioactive deoxynucleoside triphosphates by DNA polymerase for DNA sequencing. By this approach sequence information corresponding to the glucoamylase mRNA was obtained and used to determine regions of the primary structure of A. niger glucoamylase (Table III).

4. DISCUSSION

The complete amino acid sequence of glucoamylase G1 from Aspergillus niger was reconstructed from cyanogen bromide, tryptic (43), and hydroxylamine fragments in combination with cDNA sequencing data. The tryptic fragments T7-8 and T5-1 (Tca3-1) contained overlapping sequences for the cyanogen bromide fragments (43). Fourty out of the 614 amino acid residues have been deduced from the nucleotide sequence of cloned glucoamylase G1 cDNA.

Chemical analysis of CB1, the N-terminal cyanogen bromide fragment, indicated a low content of carbohydrate and O-glycosylated amino acid residues. Two positions likely to be

glycosylated are Thr(47) and Ser(93). Both were vacant in automated sequencing, but they have tentatively been identified from the amino acid composition of the highly purified and otherwise fully sequenced tryptic fragments T6-4 and T5-2, respectively (43). Ser(93) has been confirmed by the cDNA sequence (Table III). Since no other vacant positions were observed during automated sequencing of either CB1 or fragments from this region of G1, other positions can only be partially glycosylated. A short tryptic fragment, T8-1, was isolated after tryptic digestion of 2-pyridylethylated G1, CB1, or K2-CB1. Surprisingly T8-1 seemed unsusceptible to carboxypeptidases B and Y and the sequence Pro(126)-Leu(128) was established from the nucleotide sequence (Table III).

The central cyanogen bromide fragment, CB2, contained 4 glucosamine residues, suggesting the presence of two N-glycosylated units with the common pentasaccharide core (29). One site has been identified by dansyl Edman- degradation of Tca3-1 (43) to be Asn(393). Accordingly, two glucosamine residues were found in each of T5-1, Tca3-1, and a short peptide, Thr(388)-Hse(396) which was obtained by digestion of CB2 with S. aureus V8 protease. Previous experiments had indicated the content of neutral carbohydrate associated with Asn(393) to be equivalent to 8 mannosyl residues (43). The second glycosylated position in CB2 is suggested to be Asn(169). A tryptic fragment from Hy2, Asn(159)-Asn(180), contained 1.6 glucosamine residues and furthermore the requirement of an Asn-X-Ser/Thr sequence for glycosylation (26) would be fulfilled for Asn(169). The cDNA-sequence confirmed Asn at position (169) (Table III). No other vacant positions were found by automated amino acid sequencing of the region covered by CB2 and asparagine residues were not represented in the short stretches obtained by the cDNA-sequence. Thus Asn(169) and Asn(393) are suggested to be the only two fully glycosylated asparagine residues of G1. The only other potential site of N-glycosylation, Asn(180), appeared not to carry carbohydrate, since it was identified in automated peptide sequencing and the Asn(180)-Gly(181) bond was susceptible to cleavage with hydroxylamine. No residues in CB2 have been assigned as O-glycosylated. Perhaps Ser(282), which appears to be modified, carries carbohydrate (see 3.5). Indeed chemical analysis indicated the presence of O-glycosylated residues in CB2 (Table I). It is, however, assumed that they exist only as partially glycosylated positions, which would enable the determination of the amino acid sequence on the fraction of unglycosylated residues. Furthermore, it cannot be excluded that the hydroxyamino acid residues identified only by nucleotide sequencing carry carbohydrate.

The sequenced parts of Tca1-1(193-239) and W2-CB2(227-315) are overlapped by Thr(227) only. However, amino acid compositions of both fragments and the part of W2-CB2 overlapping with Tca3-2(240-271) justified the alignment. Tca3-2 and T6-3(243-271) (see reference 43) have only been sequenced to Ala(258) by peptide sequencing, while Ala(259)-Pro(267) was deduced from the nucleotide sequence (Table III). Tca1-Hy3(272-284) could only fit in Hy3 between Tca3-2 and T6-2 and the overlapping sequences were confirmed by cDNA-sequencing (Table III). Interestingly carboxypeptidase Y was observed to possess the capacity to release C-terminal 2-pyridylethyl-cysteine, from Tca3-2. T8-3(272-277) and T7-3(278-284) were found in tryptic digests of both 2-pyridylethylated G1 and CB2. The Lys(277)-Glu(278) bond seemed to be slowly hydrolyzed by trypsin, since T8-3 and T8-3+T7-3(272-284) were isolated in equimolar amounts by HPLC of tryptic digests of CB2. Presumably the adjacent Glu(278) impeded the action of trypsin at Lys(277). In T6-2, the sequence Ser(294)-Ser(296) was deduced from the cDNA (Table III). The alignment of T7-4 and T5-1 was also confirmed by nucleotide sequencing. In T7-4 the segment represented by Thr(367)-Val(375) was deduced from the cDNA sequence (Table III).

The C-terminal cyanogen bromide fragment, CB3, comprises the region characterized by a high frequency of glycosylated residues, which has recently been described in detail (43). Nucleotide sequencing confirmed that T2-dTca2-1 (498-517) follows Lys(497) as previously concluded (43); the sequence Pro(510)-Ala(512) was determined from the cDNA (Table III). The exact glycosylation pattern in the C-terminal part of the glycosylated region, Thr(498)-Thr(511), is

Table IV.
Amino acid composition of G1 and CB fragments from the amino acid sequence and from amino acid analysis.

Amino acid	Residues/fragment			Residues/G1
	CB1	CB2	CB3	
Aspartic acid	10	21	12	43
Asparagine	8 (18)	11 (31)	5 (19)	24 (67)
Threonine	12 (12)	19 (20)	42 (41)	73 (73)
Serine	15 (15)	34 (34)	35 (34)	84 (84)
Glutamic acid	4	11	11	26
Glutamine	3 (7)	10 (21)	4 (15)	17 (43)
Proline	6 (6)	7 (7)	9 (9)	22 (23)
Glycine	14 (13)	17 (17)	14 (15)	45 (45)
Alanine	15 (14)	27 (27)	21 (21)	63 (62)
Valine	8 (8)	17 (17)	15 (15)	40 (40)
Methionine	0 (0)	0 (0)	0 (0)	2 (2)
Isoleucine	7 (7)	9 (9)	7 (7)	23 (23)
Leucine	12 (12)	21 (21)	10 (11)	43 (45)
Tyrosine	4 (4)	14 (14)	9 (9)	27 (27)
Phenylalanine	3 (3)	15 (15)	3 (4)	21 (22)
Histidine	0 (0)	4 (4)	0 (0)	4 (4)
Lysine	2 (2)	5 (5)	5 (5)	12 (12)
Arginine	6 (6)	6 (6)	6 (6)	18 (18)
Homoserine	1 (1)	1 (1)	0 (0)	
2-pe-Cysteine	0 (0)	6 (5)	3 (3)	9 (8)
Tryptophan	3	8	7	18

The values in parenthesis were determined by amino acid analysis.

still unknown, because the protein sequencing was performed only on the deglycosylated peptide (43). However, from both the number of identified and the total number of O-glycosylated positions in Tca2-1, the majority of the residues in T2-dTca2-1 are suggested to be glycosylated (43). In addition CB3 was found to contain one residue of glucosamine, which has not been located in the primary structure. The absence of Asn-X-Ser/Thr sequences in CB3 excludes the possibility of a partially N-glycosylated residue. In agreement with the suggested localisation in an O-glycosidically linked oligosaccharide unit, one residue of glucosamine was lost upon HF-deglycosylation of G1. The residues in the segment Leu(519)-Gly(526) were only determined from the cDNA (Table III). The different C-terminal sequence reported earlier for G1 (44) probably reflects the difficulties encountered in interpretation of results from carboxypeptidase digestion of the long polypeptide chain.

The amino acid composition of the three CB fragments and of G1 obtained by amino acid analysis, were compared with the composition calculated from the amino acid sequence (Table IV). Only minor differences were observed. Generally the contents of 2-pe-cysteine were lower than the number of half-cystines (Tables I and IV). In the CB3 fragment the phenylalanine content was approximately 25% too high, perhaps due to the presence of a hydrolysis resistant compound coeluting with phenylalanine in the chromatogram of amino acids. The N-terminal CB fragment was low in alanine, threonine, and leucine (Table I). This probably reflects that a fraction of the G1 molecules exist in a form which lacks the three N-terminal residues (44). As can be seen from Table IV the content of asparagine in G1 was 56% of that of aspartic acid residues and the content of glutamine was 65% of the glutamic acid content. Along the polypeptide chain the frequency of the amides was rather constant

in CB1 and CB2, whereas the acids dominated in CB3. This is in agreement with G2 having a higher isoelectric point than G1 (44).

Earlier attempts to determine the molecular weight of glucoamylase G1 gave values varying from 52,000 to 110,000 (11, 23, 33, 35, 44). From the complete amino acid sequence the molecular weight of the polypeptide chain was calculated to be 65,424. The carbohydrate contents of the present preparation of G1, determined from the separated CB fragments, were 99 residues of neutral carbohydrate (mannose equivalents) plus five residues of glucosamine assumed to be N-acetylated. This corresponds to a molecular weight of the carbohydrate moiety of 17,053. Thus the molecular weight of the present preparation of glucoamylase G1 from Aspergillus niger was calculated to be 82,477[x]. In the literature, varying contents of neutral carbohydrate from 8-20% (wt/wt) and no content of hexosamine have been reported for G1 from A. niger (35). However, a few other fungal glucoamylases contain hexosamine (17, 35, 45).

No homology was apparent between the primary structures of G1 and α-amylases described in the literature (14, 21, 31, 37, 46, 47). Three fungal carbohydrases, a mycodextranase (36), a cellobiohydrolase (9, 13), and the glucoamylase G1 (34, 44), which all attack insoluble substrates, i.e. fungal cell wall polysaccharides, cellulose, and raw starch, respectively, are all glycoproteins carrying an unusually large number of short O-glycosidically linked oligosaccharide units. Recently the determination of the amino acid sequence of the cellobiohydrolase (9) was completed by aid of molecular cloning (40), but no sequence homology was found between this enzyme, which possesses a glycosylated region of about 15 amino acid residues, and glucoamylase G1. It is still not known, whether the glycosylated region is required for the formation of the insoluble enzyme substrate complexes. However, G2 does not digest raw starch, but is glycosylated to the same extent as G1 (44). Moreover it has a C-terminus corresponding to a position near the C-terminal end of the glycosylated region in G1. This suggests that other structural features are required to achieve digestion of the insoluble polysaccharides.

The glucoamylase G1 sequence presented has very recently been confirmed from residue 55 to 614 by the sequence of cDNA clones constructed from the G1 mRNA (to be published elsewhere).

Note added in proof: Recently, sequencing of cDNA synthesized complementary to the 5' end of the G1 mRNA suggests Asp in position (47) and the sequence Thr-Trp to be inserted between Tyr (50) and Thr (51) in Figure 2.

x) Recent experiments in this laboratory have reinvestigated the sedimentation equilibrium behaviour of G1. The molecular weight was now in better agreement with the calculated value. The reason for the much lower published value of 52,000 (44) is in part, that a very long period of time was needed to reach equilibrium (Ms. HANNE MØLLER and Dr. TORBEN GRAVES PEDERSEN, unpublished observations).

ACKNOWLEDGEMENTS

We are grateful to Professor MARTIN OTTESEN and Dr. NIELS P. FIIL for helpful discussions and encouragement during all stages of this work. Mss. EDITH FLØISTRUP and IBEN HJORT are thanked for outstanding technical assistance with preparation and characterization of peptide fragments and nucleotide sequencing, respectively. Mss. LONE SØRENSEN, BODIL CORNELIUSSEN, SIDSEL EHLERS and PIA BREDDAM are acknowledged for technical assistance with amino acid analysis and sequencing. Dr. ANTHONY CLARKE is thanked for linguistic correction of the manuscript.

REFERENCES

1. AVIV, H. & P. LEDER: Purification of biologically active globin messenger RNA by chromatography on oligothymidylic acid-cellulose. Proc. Natl. Acad. Sci. USA 69, 1408-1412 (1972)
2. BOEL, E., J. VUUST, F. NORRIS, K. NORRIS, A. WIND, J.F. REHFELD & K.A. MARCKER: Molecular cloning of human gastrin cDNA: Evidence for evolution of gastrin by gene duplication. Proc. Natl. Acad. Sci. USA 80, 2866-2869 (1983)
3. BORNSTEIN, P. & G. BALIAN: Cleavage at Asn-Gly bonds with hydroxylamine. Meth. Enzymol. 47, 132-145 (1977)
4. CHIRGWIN, J.M., A.E. PRZYBYLA, R.J. MCDONALD & W.J. RUTTER: Isolation of biologically active ribonucleic acid from sources en-

riched in ribonuclease. Biochemistry 18, 5294-5299 (1979)
5. DIXON, H.B.F. & R.N. PERHAM: Reversible blocking of amino groups with citraconic anhydride. Biochem. J. 109, 312-314 (1968)
6. DOWNS, F. & W. PIGMAN: Determination of O-glycosidic linkages to L-serine and L-threonine residues of glycoproteins. Meth. Carbohyd. Chem. 7, 200- 204 (1972)
7. DUBOIS, M., K.A. GILLES, J.K. HAMILTON, P.A. REBERS & F. SMITH: Colorimetric method for determination of sugars and related substances. Anal. Chem. 28, 350-356 (1956)
8. EDMAN, P. & A. BEGG: A protein sequenator. Eur. J. Biochem. 1, 80-91 (1967)
9. FÄGERSTAM, L.G.: Cellulases from Trichoderma reesei QM 9414. Enzymatic and structural properties. Dissertation, Acta Universitatis Upsaliensis no. 610, Uppsala 1981
10. FONTANA, A., D. DALZAPPO, C. GRANDI & M. ZAMBONIN: Cleavage at tryptophan with o-iodosobenzoic acid. Meth. Enzymol. 91, 311-318 (1983)
11. FREEDBERG, I.M., Y. LEVIN, C.M. KAY, W.D. MCCUBBIN & E. KATCHALSKI-KATZIR: Purification and charaterization of Aspergillus niger exo-1,4-glucosidase. Biochim. Biophys. Acta 391, 361-381 (1975)
12. FRIEDMAN, M., J.C. ZAHNLEY & J.R. WAGNER: Estimation of the disulfide content of trypsin inhibitors as S-β-(2-pyridylethyl)-L-cysteine. Anal. Biochem. 106, 27-34 (1980)
13. GUM, E.K. & R.D. BROWN: Structural characterization of a glycoprotein cellulase, 1,4-β-D-glucan cellobiohydrolase C from Trichoderma viride. Biochim. Biophys. Acta 446, 371-386 (1976)
14. HAGENBÜCHLE, O., R. BOVEY & R.A. YOUNG: Tissue-specific expression of mouse α-amylase genes: nucleotide sequence of isoenzyme mRNAs from pancreas and salivary gland. Cell 21, 179-187 (1980)
15. HAYASHIDA, S.: Selective submerged productions of three types of glucoamylases by a Black-koji mold. Agri. Biol. Chem. 39, 2093-2099 (1975)
16. HOUMARD, J. & G.R. DRAPEAU: Staphylococcal protease: A proteolytic enzyme specific for glutamoyl bonds. Proc. Natl. Acad. Sci. USA 69, 3506-3509 (1972)
17. INOKUCHI, K., T. TAKAHASHI & M. IRIE: Purification and characterization of a minor glucoamylase from Aspergillus saitoi. J. Biochem. 90, 1055- 1067 (1981)
18. JOHANSEN, J.T., C. OVERBALLE-PETERSEN, B. MARTIN, V. HASEMANN & I. SVENDSEN: The complete amino acid sequence of copper, zinc superoxide dismutase from Saccharomyces cerevisiae. Carlsberg Res. Commun. 44, 201-217 (1979)
19. JORI, G., G. GALIAZZO, A. MARZOTTO & E. SCOFFONE: Selective and reversible photo-oxidation of the methionyl residues in lysozyme. J. Biol. Chem. 243, 4272-4278 (1968)
20. KLAPPER, D.G., C.E. WILDE III & J.D. CAPRA: Automated amino acid sequence of small peptides utilizing Polybrene. Anal. Biochem. 85, 126-131 (1978)
21. KLUH, I.: Amino acid sequence of hog pancreatic α-amylase isoenzyme I. FEBS Letters 136, 231-234 (1981)
22. LAEMMLI, U.K.: Cleavage of structural proteins during assembly of the head of bacteriophage T4. Nature 227, 680-685 (1970)
23. LINEBACK, D.R. & L.A. AIRA: Structural characterization of two forms of glucoamylase from Aspergillus niger. Cereal. Chem. 49, 283-296 (1972)
24. MAHONEY, W.C. & M.A. HERMODSON: Separation of large denatured peptides by reverse phase high performance liquid chromotography. Trifluoroacetic acid as a peptide solvent. J. Biol. Chem. 255, 11199-11203 (1980)
25. MAHONEY, W.C., P.K. SMITH & M.A. HERMODSON: Fragmentation of proteins with o-iodosobenzoic acid: Chemical mechanism and identification of o-iodosobenzoic acid and a reactive contaminant that modifies tyrosyl residues. Biochemistry 20, 443-448 (1981)
26. MARSHALL, R.D.: Glycoproteins. Annu. Rev. Biochem. 41, 673-702 (1972)
27. MAXAM, A.M. & W. GILBERT: A new method for sequencing DNA. Proc. Natl. Acad. Sci. USA 74, 560-564 (1977)
28. MIAH, M.N.N. & S. UEDA: Multiplicity of glucoamylase of Aspergillus oryzae. Part 1. Separation and purification of three forms of glucoamylase. Die Stärke 29, 191-196 (1977)
29. MONTREUIL, J.: Recent data on the structure of the carbohydrate moiety of glycoproteins. Metabolic and biological implications. Pure Appl. Chem. 42, 431-477 (1975)
30. MORT, A.J. & D.T.A. LAMPORT: Anhydrous hydrogen fluoride deglycosylates glycoproteins. Anal. Biochem. 82, 289-309 (1977)
31. NAGATA, Y., S. SUGA, O. KADO & B. MARUO: N-terminal amino acid sequence of α-amylase from Bacillus subtilis var. amylosacchariticus: Comparison with that of "liquefying" type α-amylase. Agri. Biol. Chem. 44, 215-216 (1980)
32. PAZUR, J.H. & T. ANDO: The action of an

amyloglucosidase of Aspergillus niger on starch and malto-oligosaccharides. J. Biol. Chem. 234, 1966-1970 (1959)
33. PAZUR, J.H., H.R. KNULL & A. CEPURE: Glucoenzymes: Structure and properties of the two forms of glucoamylase from Aspergillus niger. Carbohyd. Res. 20, 83-96 (1971)
34. PAZUR, J.H., Y. TOMINAGA, L.S. FORSBERG & D.L. SIMPSON: Glycoenzymes: an unusual type of glycoprotein structure for a glucoamylase. Carbohyd. Res. 84, 103-114 (1980)
35. REILLY, P.J.: Starch hydrolysis with soluble and immobilized glucoamylase. Appl. Biochem. Bioeng. 2, 185-206 (1979)
36. ROSENTHAL, A.L. & J.H. NORDIN: Enzymes that hydrolyze fungal cell wall polysaccharides. The carbohydrate constitution of mycodextranase, an endo-α(1→4)-D-glucanase from Penicillium melinii. J. Biol. Chem. 256, 5295-5303 (1975)
37. ROGERS, J.C. & C. MILLIMAN: Isolation and sequence analysis of a barley α-amylase cDNA clone. J. Biol. Chem. 258, 8169-8174 (1983)
38. RUNSWICK, M.J. & J.E. WALKER: The amino acid sequence of the β-subunit of ATP synthase from bovine heart mitochondria. J. Biol. Chem. 258, 3081-3089 (1983)
39. SCHROEDER, W.A., J.B. SHELTON & J.R. SHELTON: An examination of conditions for the cleavage of polypeptide chains with cyanogen bromide: Application to catalase. Arch. Biochem. Biophys. 130, 551-556 (1969)
40. SHOEMAKER, S., V. SCHWEICKART, M. LADNER, D. GELFAND, S. KWOK, K. MYAMBO & M. INNIS: Molecular cloning of exo-cellobiohydrolase I derived from Trichoderma reesei strain L27. BIO/TECHNOLOGY 1, 691-696 (1983)
41. SVENDSEN, I., B. MARTIN & I. JONASSEN: Characteristics of Hiproly barley III. Amino acid sequences of two lysine-rich proteins. Carlsberg Res. Commun. 45, 79-85 (1980)
42. SVENSSON, B.: Covalent cross-linking of porcine small-intestine microvillar aminopeptidase. Subunit structure of the membrane-bound and the solubilized enzyme. Carlsberg Res. Commun. 44, 417-430 (1979)
43. SVENSSON, B., K. LARSEN & I. SVENDSEN: Amino acid sequence of tryptic fragments of glucoamylase G1 from Aspergillus niger. Carlsberg Res. Commun. 48, 517-527 (1983)
44. SVENSSON, B., T.G. PEDERSEN, I. SVENDSEN, T. SAKAI & M. OTTESEN: Characterization of the two forms of glucoamylase from Aspergillus niger. Carlsberg Res. Commun. 47, 55-69 (1982)
45. TAKAHASHI, T., Y. TSUCHIDA & M. IRIE: Purification and properties of three forms of glucoamylase from Rhizopus species. J. Biochem. 84, 1183-1194 (1978)
46. TÄKKINEN, K., R.F. PETTERSSON, N. KALKKINEN, I. PALVA, H. SÖDERLUND & L. KÄÄRIÄNEN: Amino acid sequence of α-amylase from Bacillus amyloliquefaciens deduced from the nucleotide sequence of the cloned gene. J. Biol. Chem. 258, 1007-1013 (1983)
47. TODA, H., K. KONDO & K. NARITA: The complete amino acid sequence of Taka-amylase A. Proc. Japan Acad. 58, Ser. B, 208-212 (1982)
48. YOSHINO, E. & S. HAYASHIDA: Enzymatic modification of glucoamylase of Aspergillus awamori var. kawachi. J. Ferment. Technol. 56, 289-295 (1978)
49. UEDA, S.: Fungal glucoamylase and raw starch digestion. Trends Biochem. Sci. 6, 89-90 (1981)

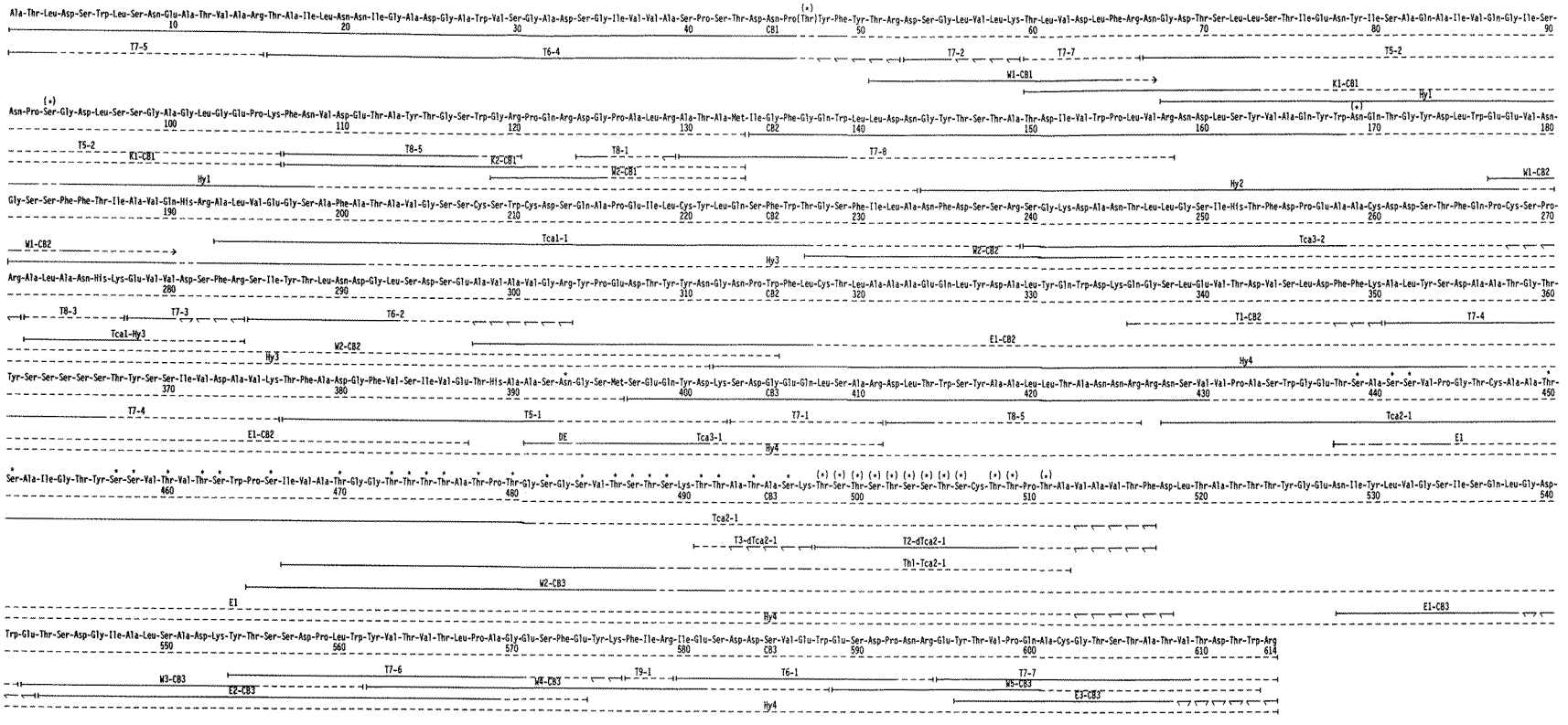

Figure 2. The complete amino acid sequence of glucoamylase G1 from Aspergillus niger. Peptides were sequenced by either automated Edman degradation (full line), digestion with carboxypeptidase B or Y (⟵), or dansyl Edman degradation (DE). (⟵) signifies identification from both the N- and the C-terminus. The full size of peptide fragments is indicated by a full line followed by a broken line under residues which were not sequenced in that particular fragment. ⟶ indicates uncertain determination of the length of a fragment. Asterisks indicate glycosylated positions. d signifies deglycosylated fragments. (*) means that the glycosylated residue has been deduced from both the finding of a vacant position by automated sequencing and the amino acid composition of the fragment. (*) means that glycosylation was assumed from chemical analysis of appropriate fragments.

Peptide fragments are designated by the following nomenclature: CB, cyanogen bromide fragments; Hy, hydroxylamine fragments; W, fragments prepared by oxidative halogenation; T, tryptic fragments; E, S. aureus V8 protease fragments; K, endoproteinase Lys-C fragments; Th, thermolytic fragments. ca, denotes fragments from citraconylated substrates. The parts of the sequence which were deduced from the nucleotide sequence of cDNA are listed in Table III.